Accidents Involving Guilty Car Drivers in Sri Lanka

DRT Jayasundara[1], CH Magalla[2]

[1]Department of Mathematics, Faculty of Engineering, University of Moratuwa, Sri Lanka
[2]Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka
[1]ravindij@yahoo.com, [2]champa@stat.cmb.ac.lk

Abstract- *Car drivers are only second to motorcycle riders in getting involved in traffic accidents in Sri Lanka. Therefore car drivers are the majority of the drivers who have been guilty for the traffic accidents in the country in recent years. The data collected by the Traffic Police, Sri Lanka on age of vehicle, age and gender of driver, ownership of the vehicle, validity and maturity of the driving license, and driver's crash factor on the accidents involving car drivers at fault (as decided by the police) happened in 2010 and 2011 are used for this study. More than 90% of car drivers who found to be liable for the accidents within this period are male, private owners of cars and more than 90% of them are holding valid driving licenses too. Thus data modeling is focused on finding the associations between age of driver, age of vehicle, maturity of driving license and driver's crash factor. The model revealed important associations between the age levels of driver, maturity of the driving license, and different reasons of drivers for the crash versus the levels of their age and the age of the vehicle.*

Keywords – traffic data, car drivers, log-linear model, odds

I. INTRODUCTION

The damages to human lives as well as public and private properties are drastically evolving due to numerous serious traffic collisions around the world. The government of Sri Lanka in line with many other countries named the first decade of Action for Road Safety 2011-2020 according to the official proclaim done at the United Nations General Assembly in March 2010 (United Nations official website). This demonstrates the high impact of traffic accidents on the society in the recent past all over the world. There have been many researches done on factors influencing the severity of traffic collisions and sometimes to predict the accident frequencies too. Negative Binomial, Poisson Regression and Multiple Linear Regression are the most popularly used statistical models in those studies. Abdel-Aty, Chen and Schott (1998) have discussed the applicability of log-linear and logit modeling with estimation of 'odds multipliers' with the intention to contribute to traffic safety studies. Even before them, log linear models have been used in accident data analyses. Even though log-linear models can evaluate many associations in a single model, in most of the research at most three variables have been modeled at a time to simplify the model. According to the analysis done by Abdel-Aty, Chen and Schott (1998), age of driver and injury severity found to have a significant association. Also they have suggested that elderly drivers might have a tendency in not driving in peak hours depending on the association discovered between average daily traffic and age of driver by their study. Olmus and Erbas (2011) have used driver's gender, whether driver is at fault or not, accident time and accident severity to determine whether there is a relationship between the variables using log linear modeling. The revealed relationships between variables by these studies contribute in many ways to understand the points pertained to traffic accidents in preventing them.

Finding the factors for traffic accidents in general would be easy but it is not easy to address them in the event of implementing safety measures. Thus, partitioning the information on traffic accidents into reachable dimensions is worthwhile in applying and assessing safety actions. According to past studies there are different types of partitioning of data used. Some of them are Truck and non-truck involved accidents (Chang and Mannering 1999), highway accidents (Jovanis and Chang ,1989), rural freeway accidents (Shankar, Mannering and Barfield ,1995), accidents at signalized intersections (Chin and Quddus,2003), accidents on multilane roads (Caliendo, Guida and Parisi ,2007). Therefore attention should be paid on analyzing the associations between variables on traffic collision data within a reasonable scope. Furthermore there is no published work found on modeling accident data in Sri Lankan context. In order to fulfill the demand for identifying the associations between variables on traffic accidents in Sri Lanka, this

research is conducted using data on age of vehicle (from the date of manufacture), age and gender of driver, ownership of the vehicle, validity and maturity of the driving license, and driver's crash factor. The analysis is streaming on modeling the variables on cars, vans, lorries, three wheelers and motorbikes. This paper discusses only the significant associations among the crucial factors for errant driving by car drivers in Sri Lanka.

II. DATA

The source of data is the official Traffic Accident Data base maintained by the Traffic Police Head Quarters, Sri Lanka. The data has been gathered using the "Road Accident Report (Form 297B)" that should be filled out by a police officer for each traffic accident reported to the police. In this analysis 8630 records on guilty car drivers are used on the accidents happened during the years 2010 and 2011. These two years are selected as they are found at the beginning of the Decade of Action for Road Safety while the number of accidents recorded in 2010 and 2011 are around the yearly average of accidents from 2003 to 2011.

III. METHODOLOGY

Cleaning and processing of raw data is done prior to the statistical analysis. The preliminary analysis is performed by examining the frequency distributions of individual variables, correlation analysis on continuous variables and chi-square testing for two-way associations of factors.

Application of log linear model for four response variables is of choice in this analysis. According to Agresti (2007) the general formulation of the saturated log linear model with four statistically dependent variables is as follows:

the general formulation of the saturated log linear model with four statistically dependent variables is as follows:

$$\log (m_{ijkh}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_h^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{ih}^{AD} + \lambda_{jk}^{BC} + \lambda_{jh}^{BD} + \lambda_{kh}^{CD} + \lambda_{ijk}^{ABC} + \lambda_{ijh}^{ABD} + \lambda_{jkh}^{BCD} + \lambda_{ikh}^{ACD} + \lambda_{ijkh}^{ABCD}$$

The log-linear model with four variables and two way interactions is formulated as follows:

$$\log (m_{ijkh}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_h^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{ih}^{AD} + \lambda_{jk}^{BC} + \lambda_{jh}^{BD} + \lambda_{kh}^{CD}$$

where;

$\log m_{ijkh}$ is the log expected frequency of cell

corresponding to level i of A, j of B, k of C and h of D.
μ is the overall effect;
λ_i^A is the effect due to the ith level of A;
λ_{ij}^{AB} is the interaction of A at the ith level and B at the jth level

Agresti (2007) illustrates the interpretations of the model which has four variables with no three-factor terms. The author specially called these models, the models with homogenous associations; where any pair of variables is conditionally associated controlling for each combination of levels of the other two variables. An absence of a two-factor term leads to conditional independence of the corresponding variables.

The calculation of odds ratios follows the following computations on log-linear parameters.

$$\frac{m_{ijkh}}{m_{1jkh}} = \exp [\lambda_i^A + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{ih}^{AD} - \lambda_1^A - \lambda_{1j}^{AB} - \lambda_{1k}^{AC} - \lambda_{1h}^{AD}]$$

Thus the odds ratio (OR) for level 1 to level 4 of B at level 1 of A and at any combination of levels of C and D is as follows:

$$\frac{m_{11kh}}{m_{14kh}} = \exp [\lambda_{11}^{AB} - \lambda_{14}^{AB}]$$

PROC CATMOD procedure in SAS 9.2 is used to fit the log-linear model. The forward selection method is used in selecting terms to the model and as a goodness of fit test, the Likelihood Ratio Statistic (G^2) is used to test the hypothesis H0: model fits the data well at 5% significance level.

Even though a goodness-of-fit test is usually in use for finalizing a model it should not be the one and only criterion (Agresti 2007). Alternatively, the dissimilarity index is introduced as a measure of the closeness of a model fit when the sample size is large. For the cell counts in any higher dimensional table (m_i) and fitted values(e_i) the dissimilarity index is calculated by

$$D = \sum |m_i - e_i|/2n;$$

where n is the sample size. This index varies between 0 and 1. Smaller values represent a better fit. The value of D represents the proportion of sample cases that must move to different cells for the model to achieve a perfect fit. Therefore having a very small D value indicates that the sample data follow the model pattern closely, even though the model is not proved by the G^2 test. Thereby the dissimilarity index helps indicate whether the lack of

fit is important in a practical sense (Agresti, 2007).

VII. STATISTICAL ANALYSIS AND INTERPRETATION

According to the frequency of accidents having a car driver being guilty for the collision, more than 90% of the cars involved in accidents during 2010-2011 were private vehicles with male drivers who were holding a valid driving license. Thus only age of vehicle, age of driver, maturity of the driving license, and driver's pre-crash factor were tested for two-way associations using chi-square testing. The chi-square statistics resulted from these tests were significant for every pair of variables. Therefore, the data on the same variables has been considered for model fitting. The levels of the variables are as follows:

A =Age of vehicle (vehiage)
A_i = i^{th} category of vehiage
i=1: less than 7 years
i=2: 8 to 15 years
i=3: 16 to 20 years
i=4: above 20 years
B =Age of driver (age)
B_j = j^{th} category of age
j=1: 18 to 24 years
j=2: 25 to 44 years
j=3: 45 to 64 years
j=4: above 64 years
C =Maturity of the driving license (maturity)
C_k = k^{th} category of maturity
k=1: less than 1 year
k=2: 2 to 5 years
k=3: above 5 years

D =Driver's pre-crash factor (DPF)
D_h = h^{th} category of DPF
h=1: speeding
h=2: aggressive/negligent driving
h=3: error of judgment
h=4: influenced by alcohol/drugs
h=5:distracted/inattentiveness due to fatigue, fall asleep, handling mobile phone/radio, etc.

Model Fitting

Table 1: Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr> ChiSq
age	3	1022.56	<.0001
maturity	2	78.79	<.0001
DPF	4	3375.04	<.0001
vehiage	3	754.29	<.0001
vehiage*DPF	12	100.42	<.0001
maturity*DPF	8	47.21	<.0001
vehiage*age	9	41.79	<.0001
age*maturity	6	1077.06	<.0001
G^2 D=0.0377	192	182.99	0.6675

Table 1 shows the results of the chi-square tests performed by SAS package for the significance of terms in the model. The test of the goodness of fit of the fitted model with H0: model fits the data well versus H1: model does not fit the data well, resulted a Likelihood Ratio Statistic (G^2) = 182.99 with DF= 192. The p-value of G^2 is 0.6675 which greater than 0.05. Thus the statistical evidence given by the test is strong enough to not to reject the null hypothesis. Further the dissimilarity index is also calculated and it is 3.77%. This confirms concluding the model as a close fit to the data as it says only approximately 4% of sample data must move to different cell to achieve the best fit. Therefore using both criterions above, the model,

$$\log (m_{ijkh}) = \mu + \lambda_i^{vehiage} + \lambda_j^{age} + \lambda_k^{maturity} + \lambda_h^{DPF} + \lambda_{ij}^{vehiage*age} + \lambda_{ih}^{vehiage*DPF} + \lambda_{jk}^{age*maturity} + \lambda_{kh}^{maturity*DPF}$$

is finalized and is used to describe the associations between variables.

A. Important associations between driver's pre-crash factor and age of vehicle in car accidents

Table 2: Odds Ratios for parameter estimates (with p<0.05) of log linear model for interaction between DPF and vehiage.

Driver's Pre-crash Factor	Compared levels of vehicle age	Odds Ratio (OR)
Error of Judgment	< 7 years to > 20 years	2.3
Influenced by alcohol/drugs	> 20 years to < 7 years	3.5

The odds of a car crash due to errant judgment by driver for cars not older than 7 years is more than two times higher compared to cars older than 20 years . At the same time data shows no significant association between the driver's pre-crash factor and the age of the driver. Therefore it can be suggested that the over confidence in drivers on their new cars can cause this errant judgment about the fore coming instance which has led to an accident.

The next most associated factor with the age of vehicle is the influence of alcohol or drugs. The odds of a car crash due to a driver influenced by alcohol or drugs for cars which are older than 20 years is 3.5 times higher than that for the cars not older than 7 years. In the situations when the driver is not conscious because of the influence of alcohol or drugs, even though they take sudden actions to prevent accidents the expected quick response from the car may not be available through the condition of an old car as from a new one. Also the drivers may not be that serious about their cars when the car is old.

Important associations between driver's pre-crash factor and maturity of license in car accidents

Table 3: Odds Ratios for parameter estimates (with p<0.05) of log linear model for interaction between DPF and maturity.

Driver's Pre-crash Factor	Compared levels of license maturity	Odds Ratio (OR)
Influenced by alcohol/drugs	< 1 year to >5 years	2.03

The interaction between driver's pre-crash factor and the maturity of license became significant in the model. The model coefficient of the effect of "influenced by alcohol/drugs when license maturity less than 1 year" is highly significant at 0.05 level of statistical significance. Thus it reveals that the odds of accidents due to a car driver influenced by alcohol or drugs is 2 times higher in drivers having a driving license which is not matured more than an year with reference to the drivers having a license matured more than 5 years. Thus it is clearly a red alarm on new drivers not to drink and drive.

B. *Important associations between age of driver and age of vehicle in car accidents*

Table 4: Odds Ratios for parameter estimates (with p<0.05) of log linear model for interaction between age and vehiage.

Age of Driver	Compared levels of vehicle age	Odds Ratio (OR)
45 to 64 years	< 7 years to > 20 years	1.16
25 to 44 years	< 7 years to > 20 years	1.21
25 to 44 years	8 - 15 years to > 20 years	1.21

These odds ratios are not much different from 1. But these odds ratios have been evident by a significant number of accidents in comparison to other levels of the interaction between age of driver and age of vehicle. Thus they are statistically significant at the 5% significance level. Consequently, the model coefficient at the interaction level of 45 to 64 years of age of driver and vehicle age less than 7 years is statistically highly significant at 0.05 level. Therefore it can be concluded with enough statistical evidence that the odds of having the car driver's age between 45 and 64 at an accident is equally likely for cars which are not older than 7 years compared for cars which are older than 20 years.

Further, for the car drivers within the age group 25 to 44, the odds of getting involved in accidents when the car is not older than 15 years is as equal as when the car is older than 20 years. The reason can be the over confidence in driver on his driving ability and the condition of the car. Since the three-way interaction between age of driver, age of vehicle and the deriver's pre-crash factor is not significant in the model, it can be said that, in these situations there is a possibility of any pre-crash factor at the accident circumstances.

C. *Important associations between age of driver and maturity of license in car accidents*

There is a moderate positive correlation between age of driver and maturity of license which has a Pearson's correlation coefficient 0.648. This reveals that even among the car drivers who have been guilty for accidents, when age of the driver increases there is an increase in the maturity of their license. Therefore the odds ratios are not calculated for the levels of age of driver versus maturity of license.

V. DISCUSSION

The fitted log-linear model can vary depending on the categorizations done in continuous variables. The categorization of Age of vehicle is done based on the categorization used by Sri Lanka Insurance in providing insurance facilities at different conditions. The margin of the final level of Maturity of License is set to five years considering the license renewal policies used by the Department of Motor Traffic, Sri Lanka. The categorization of Age of Driver is done in consultation with the officers in IT Division of the Head Quarters of Traffic Police Sri Lanka in line with a categorization used in a study of Norman, 1962.

VI. CONCLUTIONS

According to traffic accident data from 2010 and 2011 by Traffic Police in Sri Lanka, most of the blameworthy car drivers were male, private car owners who held valid driving licenses. The most crucial situations revealed through them are driving old cars by drunk drivers and driving cars by drunk drivers who are with a license not older than 1 year. Furthermore, there is a tendency of making more car accidents by the drivers who drive cars which are not older than 7 years due to making wrong decisions, in comparison to the car drivers who drive older cars.

REFERENCES

Ciro Caliendo , Maurizio Guida , Alessandra Parisi (2007), A crash-prediction model for multilane roads, Accident Analysis and Prevention Vol.39 657–670

Hoong Chor Chin, Mohammed Abdul Quddus (2003), Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections, Accident Analysis and Prevention 35 253–259

H. Olmuş, S. Erbaş: Analysis of Traffic Accidents Caused by Drivers by Using Log-Linear Models(2011), Traffic & Transportation, Vol. 24, 2012, No. 6, 495-504

Li-Yen Chang, Fred Mannering (1999), Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents, Accident Analysis and Prevention Vol.31, 579–592

Mohamed A. Abdel-Aty, Chien L. Chen and James R. Schott (1998), An Assessment of the Effect of Driver Age on Traffic Accident Involvement Using Log-Linear Model, Accident Analysis and Prevention, Vol. 30, No. 6, pp. 851–861

Norman L.G(1962), Road Traffic Accidents: Epidemiology, Control and Prevention, World Health Organization, Geneva, Printed in New Zealand, http://whqlibdoc.who.int/php/WHO_PHP_12.pdf (Accessed on 05/01/2014)

Prabhani Kuruppumullage and Roshini Sooriyarachchi (2007), Log-linear models for ordinal multidimensional categorical data, National Science Foundation Sri Lanka, 35(1) 29-40

Paul P. Jovanis, Hsin-Li Chang (1989), Disaggregate Model of Highway Accident Occurrence Using Survival Theory, Accident Analysis and Prevention Vol. 21, No. 5, pp. 445-458

United Nations Decade of Action for Road Safety 2011-2020 http://www.un.org/en/roadsafety [Last access on 22/05/2014]

Venkataraman Shankar, Fred Mannering, and Woodrow Barfield (1995), Effect Of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies, Accident Analysis and Prevention, Vol. 27, No. 3, pp. 371-389

www.dbeBooks.com - An Ebook Library Alen Agresti (2007) An Introduction to Categorical Data Analysis. Wiley, New Jersey, ISBN 978-0-471-22618-5

BIOGRAPHY OF AUTHORS

D.R.T. Jayasundara is a lecturer in Statistics and Operational Research attached to the Department of Mathematics, Faculty of Engineering of University of Moratuwa, Sri Lanka.

Dr. C. H. Magalla is a senior lecturer in Statistics attached to the Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka.